Ontology-based Land Use Information Service on the Semantic Web

Yuxia Huang

Department of Geography, University at Buffalo, Buffalo, NY, 14261

E-mail: huang3@acsu.buffalo.edu

Abstract

With billions of Web pages on the World-Wide-Web (WWW) and millions of users, how to access, exchange and integrate spatial information is one of the keys to the WWW success. However, today's web is designed for direct use by people, who must view, understand, select and navigate its information. Moreover, it is difficult to use because of the semantic conflict between data and lack of tools to allow integrated access to shared spatial information. In this paper, we present an approach by using the semantic web technology and the common ontology to help people find spatial information in an Internet-connected world. First, we introduce the semantic web, including its main concepts, such as RDF, ontology, and agents. Second, we describe Semantic Conflict Resolution Ontology (SCROL), which proposed by Ram and Park, to provide a systemic method for automatically detecting and resolving semantic conflicts. Third, we show how to make land use information service work on the semantic web. Finally, we have the conclusion and discuss the future work.

Keywords

Semantic web, ontology, land use information service, semantic interoperability

1 Introduction

For the past few years, World-Wide-Web (WWW) has developed dramatically fast and has drastically changed the availablity and access the information. People can easily get a lot of information by typing the key words using search agents, such as google, yahoo. But with billions of Web pages on the World-Wide-Web (WWW) and millions of users, how to access, exchange and integrate spatial information is one of the challenges to the WWW success. Currently, although we can get a lot of information from the WWW, finding the right piece of information is often a nightmare(Fensel, Hendler et al., 2002).That is because today's web is designed for direct human readable and human consumption. People must view, understand, select and navigate the information. Therefore, after inputting the request which the users want to search, the computers find tons of information, then the users have to go through all the information and finally get the information by themselves what they want. Moreover, it is difficult to use the search result because of the semantic conflict between data. Semantic conflicts occur whenever two contexts do not use the same interpretation of the information (Wache and Vogele, 2001). Goh (Goh, 1997) presented three main causes of semantic conflict: *confounding conflicts* occur when information items have the same meaning, but different in reality; *scaling conflicts* occurs when the same value was measured by the different reference systems, and *naming conflicts* occurs when name rules are different in different system, such as homonyms and synonyms. Furthermore, the lack of tools to allow integrated access to shared spatial information. Usually the users search result can not just come directly from the piece of information, but from the process and integrate from the information resources.

Semantic web, proposed by Tim Berners-Lee, is the new technology to share common formalized knowledge via the Internet. The central idea in Semantic Web is to develop and use machine-understandable instead of human-understandable language for the expression of the semantic content of Web pages, to make the Web more intelligent, therefore, to help people find the right information on the Web.

Key issue for sharing knowledge is the use of a common terminology, e.g, in form of ontologies (Fensel, Hendler et al., 2002). Understanding the meaning and the use of terminology from different domains and the mapping ability between agreed concepts are important to make a semantically compatible spatial information environment. One of the necessary issues for achieving it concerns the identification of semantically related information in difference sources and the subsequent resolution of the differences among the information. There are several techniques to determine semantically related information, such as semantic modeling approaches, formal logic-based approach,knowledge-based system, and the use of a common ontology.

In this paper, we present an approach to find spatial information based on the use of common ontology and service agents by using Semantic Web technology. The success of approach will lead to knowledge acquirement from spatial information on the WWW in the future. The remainder of the paper is structured as follows: In section 2, we discuss the semantic web. In section 3, we describe Semantic Conflict Resolution Ontology (SCROL), which proposed by Ram and Park (Ram and Park, 2001), to provide a systemic method for automatically detecting and resolving semantic conflicts. In Section 4, we show how to make land use information service work on the semantic web. Finally, we have the conclusion and discuss the future work.

2. The Semantic Web

The semantic web has been presented as the next step in the evolution of the WWW (Berners-Lee, 1999). By focusing on annotating data using ontologies to make data machine-understandable, the semantic web will bring structure to the meaning of Web pages, to build an semantic-level environment so that the software agents can roam from page to page to carry out the sophisticated tasks for users.

According to Berners-Lee(Berners-Lee, 1999), achieving a semantic web requires:

- Developing languages and terminologies, where languages are used to express the machine-understandable meta-information for the documents, and terminologies such as ontologies are used to make the information languages expressed is semantic consistence.

- Developing tools and new architecture that use such languages and terminologies to access, exchange and integrate information.

- Realizing applications that provide a new level of service to the human users of the semantic web.

2.1 Languages for the semantic Web

Languages for the Semantic Web must include two aspects (Fensel, Hendler et al., 2002): first, formal syntax and formal semantics to annotate data to make the information is machine-understandable, therefore, the software agents can process automatically the request. Second, provide standardized vocabulary (e.g. ontologies) to make human agents can share the information and knowledge. Figure 1 shows the languages for the Semantic Web.

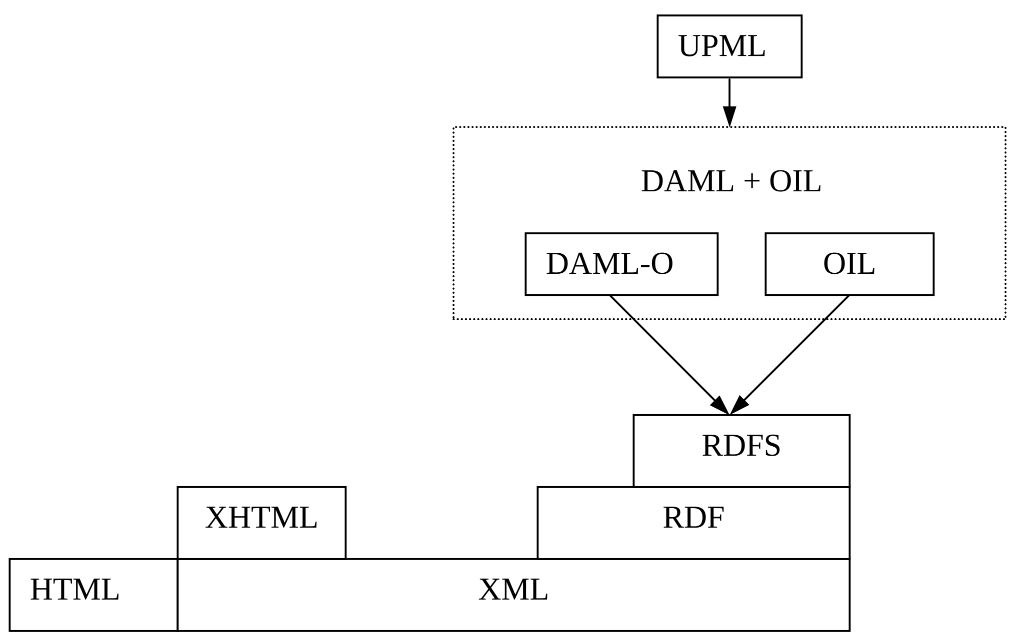

Figure 1 Layer Language model for the WWW (Fensel, Hendler et al., 2002)

2.2 Resource Description Framework (RDF)

As the formal language, Resource Description Framework (RDF), which based on the existing standards of XML, URIs and Unicode, defines a syntactical convention and a simple data model for representing machine-processible semantics of data. RDF is a foundation for processing metadata (Miller 1998), where metadata on the WWW can be encoded, exchanged and reused. RDF consists of three object types:

- Resources: resources may be part of a Web page, an entire Web page, a collection of pages, or an object that is not directly accessible via the Web. Resources are always named by URLs.

- Properties: A property is a characteristic, attribute or relation used to describe a resource.

- Statements: Each RDF statement has three parts: a resource, a named property together with the value of the property for that resource.

RDF does not stipulate semantics for each resource, but rather make statements that are machine-processable, therefore, to provide a means for software agents to exchange the information on the Web.

2.3 Ontology

RDF can the information became machine-processable, but how to make the machine understand and share the meaning of the information? The concept of the ontology, which is the central to the Semantic Web, can be developed to solve these problems(Swartz). Since the beginning of the 1990s, ontologies, which provide a shared and common understanding of a domain that can be communicated between people and heterogeneous application system, have become a popular topic in artificial intelligence, including knowledge engineering, natural-language processing, and knowledge presentation (Fensel, Hendler et al., 2002). An ontology is a formal, explicit specification of a shared conceptualization of a certain domain of discourse in a computer system (Gruber, 1995; Fensel and van Harmelen, 2001). Where, conceptualization refers to an abstract, simplified view of the world that identifies that phenomenon's relevant concepts. Explicit means that explicitly defined the type of concepts and constraints used, and formal means that the ontology is machine- understandable. Many ontology languages can be used in the Semantic Web, such as DARPA Agent Markup Language with Ontology Inference Layer (DAML + OIL) can be used to build an ontology vocabulary.

Usually there are two ways to resolve semantic conflict using the concept of ontology. One way is to build the ontology for each system, then directly create the mapping between these ontology-based systems. Another way is to create common ontology to specify the vocabularies and terms to describe and interpret shared information among its users, and build the mapping between the systems through the common ontology. This paper we will use the common ontology to facilitate the semantic-level information service.

2.4 Agents

Frauenfelder (Frauenfelder 2001) concluded four basic components necessary for building the Semantic Web, such as machine readable, semantic indexing, semantic services and reasoning. Agents, which are the software components, play an important role. Many kinds of agents are needed to be defined and be accomplished to automatically retrieve and process information from various sources and fulfill queries. Agents not only improve the user experience with regards to keeping information organized, but also perform tedious tasks or well-defined processes for the user.

2.5 Semantic Conflict Resolution Ontology (SCROL)

In this paper, Semantic Conflict Resolution Ontology (SCROL), which proposed by Ram and Park (Ram and Park, 2001), will be adopted to describe and organize the semantic of the common standard vocabulary and the relationship between them, including concepts, instances, relationship between concepts, and relationship between instances.

SCROL is a formally defined ontology, which can be used to automatically identify and resolve semantic conflicts among heterogeneous databases. A tree structure is used in SCROL to express the semantic of each information source. Therefore, SCROL is a tuple of Ω = (OC, OI, RS, RM, μ), where:

OC: is a distinct set of concepts. Concepts are represented as terms. Concepts have properties, such as name, definition, subconcept-of, instance, and referenced-by.

OI: is a distinct set of instances. Instances are also represented as terms. Instances have properties including name, definition, instance-of, referenced-by.

RS: refers to a sibling relationship and is a relation on OC. RS can occur only between two concepts, but not between two instances. RS consists of four kinds of relationship: disjoint, peer, part-of, and is-a relationship. Where, disjoint indicates means not semantically equivalent. Peer means two concepts are semantically equivalent. Part-of is similar to an “aggregation” . And is-a is a generalization/specialization.

RM: a relation on OI, and called a domain value mapping relationship. RM can occur only between instances, but not between concepts. RM consists of: one-one, one-many, many-many, or none.

μ : the root of Ω . The rood μ has no parent, and there is exactly one μ in Ω .

3. Semantic Web-based Land Use Information Service

In this paper, in order to understand the meaning and the use of terminology from different domains and have the mapping ability between agreed concepts to solve semantic conflict, we use the common ontology approach to acquire information regarding the meaning of terms, and to compare two terms to determine if and how they might be related. First, the agreements on the meaning of the terms used in a common ontology will be established. Second, the formal structure of Semantic Conflict Resolution Ontology (SCROL) will be used to describe, interpret and organize the semantic of the common standard vocabulary and the relationship between them, including concepts, instances, relationship between concepts, and relationship between instances. Third, the type of semantic conflict and the set of rules to detect and resolve the semantic conflicts will be established. Finally, the mapping between information sources and the common ontology will be established. In addition, the ontology will be presented using DAML+OIL language and the resources will be presented by RDF to make agents can reach the resources on the WWW. Service agents are also provided to help people find, process, and integrate spatial information. There are three main types of service agent in this paper: 1) task analysis agent, the responsibility is to understand the request by the people, and decompose the request into several sub-tasks; 2) search agent, the responsibility is to capture and discovery the information on the WWW; 3) information process agent, the responsibility includes detect and resolve the semantic conflict between the information, spatial information process, such as projection transform, overlay, buffer, join, and integrate the spatial information.

To see how spatial information service works on the semantic web, we provide a simple example of a user trying to find land use information. Imaging a user is looking for SUNR county land use information in 2000 on the WWW, and from the common ontology, we know SUNR county is composed of three cities: Maro, Neok, and Gare. Let’s assume that no SUNR land use information in 2000 is available on the WWW, instead of, these three cities’ land use information in 2000 are available on the WWW, including shape file and metadata. All of them use the same scale (1:250,000), However, each city’s land use information data and Neok use the same land use category, but Gare uses the different land use category (Table 1 shows the comparison of these cities’ land use category). Regarding to the coordinate system they used, Maro data cast to the Universal Transerve Mercator (UTM) projection, and referenced to the North American

Datum of 1983 (NAD83), but Neok and Gare land use data use geographic coordinate system.

In order to get the result from the WWW, first, the common ontology for land use category will be established and organized by SCROL, which the first level includes residential, commercial, service and institutional, industrial, transportation, communications and utilities, mixed land so on. Due to the space, only the part of residential and commercial, service and institutional are expanded in figure 2. Then, the semantic web will be built, where, information about providing Maro, Neok and Gare land use information will be published on the WWW by XML (eXtensible Markup Language), and the related metadata such as coordination system and land use category will be described by RDF (Resource Definition Framework). When the user sends the request like "help me find SUNR county land use information in 2000 in UTM coordination system." Then, the mapping between the land use category in SCROL and the land use category used and coordinate system in each city will be built. Each component in local schema will mapped to either a concept or instance in SCROL. For example, the "UTM" in local schema 1 is mapped to "UTM" in SCROL, "Geographic Coordinate System" is mapped to "Geographic Coordinate System" in SCROL. Because the relationship between "UTM" and "Geographic Coordinate System" is "peer", it means they can be converted each other. Figure 2 shows the part of content in SCROL, local schema for each city, and the mapping between local schema and SCROL.

Then, the main search and process procedures are: 1) Based on the knowledge that SUNR county includes three cities: Maro, Neok and Gare, task analysis agent decompose this request into three sub-task: find Maro land use information in 2000, find Neok land use information in 2000, and Gare land use information in 2000. 2) Search agent finds these cities' land use information. 3) Get the coordination system for each information. If it is not UTM, projection transform agent will transform its coordination system to UTM. 4) Get the land use category for each information. Semantic process agent will detect the category conflict between the land use category and the common land use ontology based on the mapping between them, and convert each city's land use category to the common ontology. For example, in city Maro's land use data, "office" item is mapped to the "office" in the common ontology, but "office" item and "retail and other commercial" will be combined to get "commercial" information in the common ontology. 5) Finally, information integrate agent will join these new three city's land use information and get SUNR county land use information, then send it to the user.

Table 1 The comparison of land use category between city Maro, Neok and Gare

Land use category used in the city Maro and Neok	Land use category used in the city Gare
Single family residential Multifamily residential Office Retail and other commercial Mixed use Industrial and utility Extractive Institutional Park, recreational and preserve Golf course highway railway airport agriculture undeveloped water	Single-family residential Multi-family residential, high rise Multi-family residential, low rise Commercial Industrial Institutional establishments Airport Parks/recreation agricultural Highways water Undeveloped industrial parks Undeveloped public land Mixed use residential Mixed use industrial Mixed use commercial and other

4. Conclusion and future research

The semantic web has been presented as the next step in the evolution of the WWW. In this paper, we present land use information service using the Semantic Web technology to access the right information. SCROL is used to manage and organize the common ontology. This is a primitive work. The relative work are undergoing.

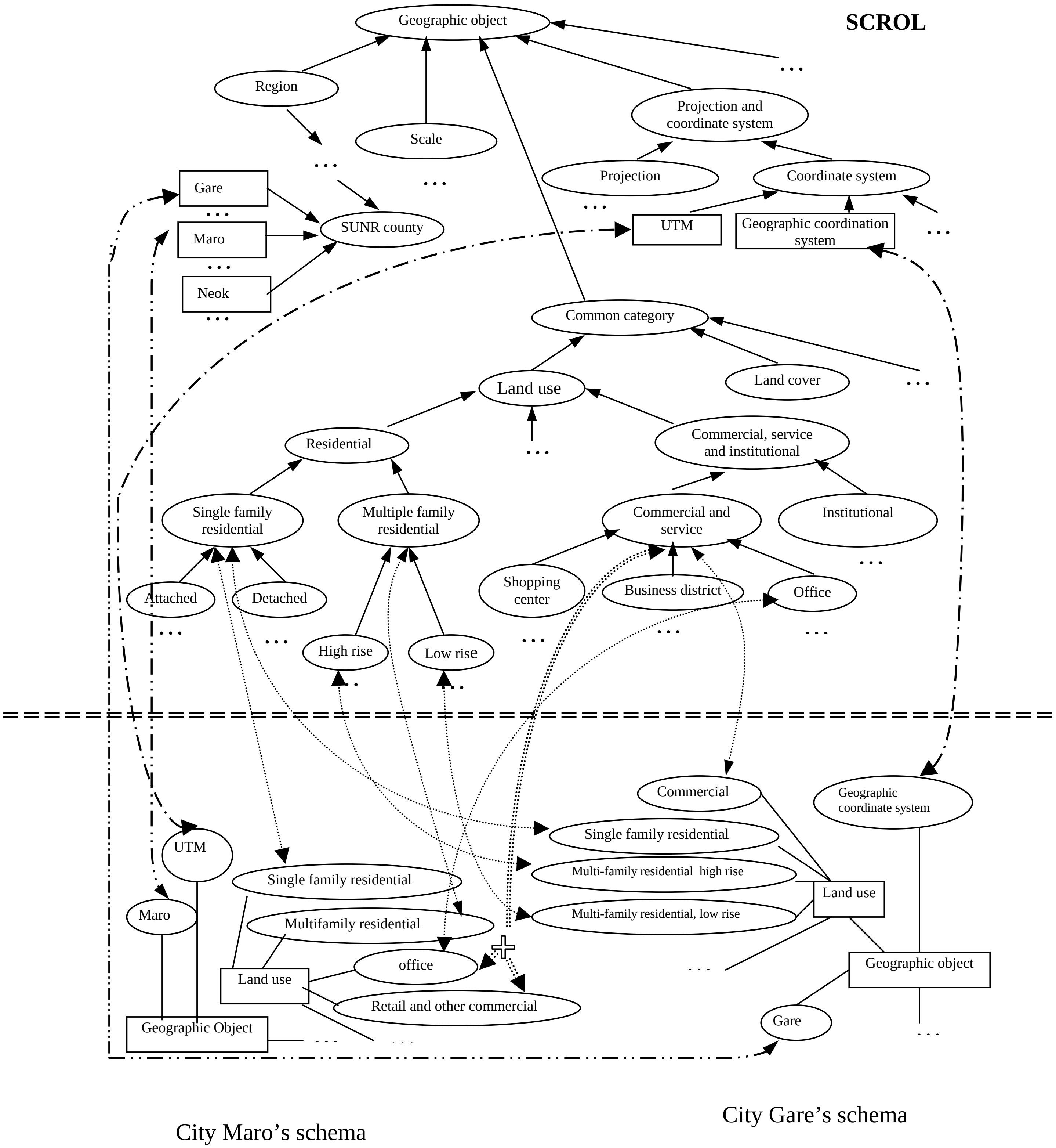

Figure 2 An example of SCROL , local schema and Ontology mapping

Acknowledgements

I would like to thank Dr. Ling Bian for her suggestion.

References

Berners-Lee, T., 1999, Weaving the Web, the original design and ultimate destiny of the World Wide Web, Harper

Casey, M., and Austin, M. A., 2001, Semantic web methodologies for spatial decision support, DISAge2002

Fensel D., van Harmelen F., 2001, OIL: an ontology infrastructure for the semantic web, IEEE Intelligent Systems, March/April, pp.38-45

Fensel D., Hendler, J., Lieberman, H., and Wahlster W., 2002, Spinning the Semantic Web, The MIT Press, Cambridge, Massachusetts, London, England

Frauenfelder, M., 2001, A Smarter Web, MIT technology review, November, 2001

Goh, Ch., 1997, Representing and reasoning about semantic conflicts in heterogeneous information sources, PhD MIT

Gruber, Th.R., 1995, Toward principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies, 43(5-6), pp.907-928

Miller, E., 1998, An introduction to the resource description framework, D-Lib magazine, May.

Ram, S., and Park, J., 2001, Semantic conflict resolution ontology(SCROL): An Ontology for detecting and resolving data- and schema-level semantic conflict, working paper

Wache, H., Vogele, T., Visser, U, Stuckenschmidt, H., Schuster, G., Neumann, H., and Hubner, S., 2001, Ontology-based integration of information – A survey of existing approaches, Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, Vol. Pp. 108-117

Swartz, A., The semantic web in breadth, http://logicerror.com/semantic web-long